Forecasting bubbles with mixed causal-noncausal autoregressive models

Voisin, Elisa and Hecq, Alain

Maastricht University

13 March 2019

Forecasting bubbles with mixed causal-noncausal autoregressive models

Alain Hecq and Elisa Voisin [1]

Maastricht University

August, 2019

Abstract

This paper investigates one-step ahead density forecasts of mixed causal-noncausal models. It analyses and compares two data-driven approaches. The paper focuses on explosive episodes and therefore on predicting turning points of bubbles. Guidance in using these approximation methods are presented with the suggestion of using both of the approaches as they jointly carry more information. The analysis is illustrated with an application on Nickel prices.

Keywords: noncausal models, forecasting, predictive densities, bubbles, simulations-based forecasts.
JEL. C22, C53

[1]Corresponding author : Elisa Voisin, Maastricht University, Department of Quantitative Economics, School of Business and Economics, P.O.box 616, 6200 MD, Maastricht, The Netherlands. Email: e.voisin@maastrichtuniversity.nl.
The authors would like to thank Sean Telg , Sebastien Fries, Eric Beutner, two anonymous referees as well as the participants of the CFE 2018, Pisa, for valuable comments and suggestions. All remaining errors are ours.

1 Introduction

Locally explosive episodes have long been observed in financial and economic time series. Such patterns, often observed in stock prices, can be triggered by anticipation or speculation. Given this forward-looking aspect, expectation models have been prevalent for modelling them. As shown for instance by Gouriéroux, Jasiak, and Monfort (2016), equilibrium rational expectation models admit a multiplicity of solutions, and some of them feature such speculative bubble patterns.[2] Models employed to capture them range from simplistic approaches, such as single bubble models with constant probability of crash, to rather complex models depending on numerous parameters. Although those models may a posteriori fit the data well, they are either not informative enough or render predictions uncertain due to their dependence on extensive parameters estimation.

This paper analyses and compares two data-driven approaches to perform density forecasts of mixed causal-noncausal autoregressive (hereafter MAR) models. MAR models incorporate both lags and leads of the dependent variable with potentially heavy-tailed errors. The most commonly used distributions for such models in the literature are the Cauchy and Student's t-distributions. While being parsimonious, MAR models generate non-linear dynamics such as locally explosive episodes in a strictly stationary setting (Fries and Zakoïan, 2019). So far, the focus has mainly been put on identification and estimation. Hecq, Lieb, and Telg (2016), Hencic and Gouriéroux (2015) and Lanne, Luoto, and Saikkonen (2012) show that model selection criteria favour the inclusion of noncausal components explaining respectively the observed bubbles in the demand of solar panels in Belgium, in Bitcoin prices and in inflation series. Few papers look at the forecasting aspects. Gouriéroux and Zakoïan (2017) derive theoretical point and density forecasts of purely noncausal $MAR(0,1)$ processes with Cauchy-distributed errors, for which the causal conditional distribution admits closed-form expressions. With some other distributions however, like Student's t, conditional moments and distribution may not admit closed-form expressions. Lanne, Luoto, and Saikkonen (2012) and Gouriéroux and Jasiak (2016) developed data-driven estimators to approximate them based

[2] In this paper, speculative bubbles, or simply bubbles are referred to as processes characterised by a rapid and persistent increase followed by a crash. Some authors talk about bubbles to denote the deviation from the fundamental solution of a present value type model. Those bubbles might not have the non-linear pattern that we investigate in this paper.

on simulations or on past realised values respectively, applicable to any distribution. Nonetheless, the literature regarding the ability of MAR models to predict both explosive and stable episodes remains scarce (see also Lanne, Nyberg, and Saarinen, 2012 and Gouriéroux, Hencic, and Jasiak, 2018). The aim of this paper is to analyse and compare in details the two numerical methods mentioned for forecasting $MAR(r,1)$ models, with unconstrained r number of lags, a unique lead and a positive lead coefficient. Furthermore, the focus is put on positive bubbles since they are prevalent in financial and economic time series. This paper investigates the possibility to predict, one-step ahead, probabilities of turning points of locally explosive episodes. We find that the sample-based method is characterised by a learning mechanism and that the simulations-based approach is a good approximation of theoretical results. Our results show that combining results obtained from the two methods can help disentangling the proportion of probabilities induced by the underlying distribution and by past behaviours. This information could for instance be used for investment decisions, where the strategy is to be built based on the investor's risk aversion and beliefs regarding the series.

The paper is constructed as follows. Section 2 introduces mixed causal-noncausal autoregressive models. Section 3 discusses how they have been used for prediction so far when the conditional moments and densities admit closed-form expressions. In Section 4 are presented the simulations-based forecasting approach proposed by Lanne, Luoto, and Saikkonen (2012), followed by the sample-based method proposed by Gouriéroux and Jasiak (2016). The performance of both approaches is compared, when available, to theoretical results. The analysis is based on various $MAR(0,1)$ processes with Cauchy or Student's t-distributed errors. In Section 5 both approximation methods are illustrated with an application to a detrended Nickel prices series. Section 6 concludes.

2 Mixed causal-noncausal autoregressive models

Consider the univariate $MAR(r,s)$ process defined as follows,

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,$$

where L and L^{-1} are respectively the lag and forward operators; Φ and Ψ are two invertible polynomials of degree r and s respectively. That is, $\Phi(L) = (1 - \phi_1 L \cdots - \phi_r L^r)$ and $\Psi(L^{-1}) = (1 - \psi_1 L^{-1} \cdots - \psi_s L^{-s})$ with roots strictly outside the unit circle. The error term ε_t is i.i.d, following

a non-Gaussian distribution. This assumption, not empirically restrictive since non-normality is widely observed in financial and economic time series, is necessary to achieve identification of the model. An $MAR(r,s)$ model can also be expressed as a causal AR model where y_t depends on its own past and present value of u_t,

$$\Phi(L)y_t = u_t, \tag{1}$$

where u_t is the purely noncausal component of the errors, depending on its own future and on the present value of the error term

$$\Psi(L^{-1})u_t = \varepsilon_t. \tag{2}$$

Alternatively, we can also filter the process as $\Phi(L)v_t = \varepsilon_t$ with $\Psi(L^{-1})y_t = v_t$ to obtain the backward component of the errors, v_t. The process y_t admits a stationary infinite two-sided MA representation and depends on past, present and future values of ε_t,

$$y_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}.$$

The case in which all coefficients a_i for $-\infty < i \leq 0$ (resp. $0 \leq i < \infty$) are equal to 0, corresponds to a purely causal (resp. noncausal) model.

Despite their apparent simplicity and parsimony, MAR models often provide a better fit to economic and financial data as they capture non-linear causal dynamics such as bubbles or asymmetric cycles. The shape of series generated by $MAR(r,s)$ processes depends on the presence of leads, lags and the magnitude of their coefficients. Figure 1 displays how the presence of a lag, a lead, or both, affects the shape of transitory shocks in MAR series. Purely causal (resp. noncausal) processes are only affected by a shock after (resp. before) the impact; this is shown in graph (a) (resp. (b)). Consequently, MAR processes are affected both in anticipation and after the shock; the shape of the explosive episode (mostly forward or backward looking) depends on the magnitude of the lag and lead coefficients. When the coefficients are identical (c) the effects of the shock are symmetric around the impact while when the coefficient of the lead is higher (d), the explosive episode is more analogous to what we refer to as a bubble with an asymmetry around the peak.

The usual practice for estimating and identifying MAR models is as follows. Methods based on first and second moments (e.g. OLS) are unable

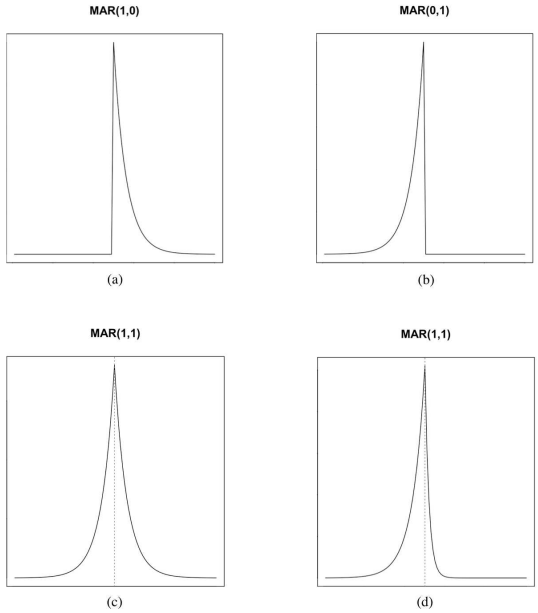


Figure 1: Effects of a lag and a lead on transitory shocks for MAR series (a) purely causal $\phi = 0.8$ and $\psi = 0$, (b) purely noncausal $\phi = 0$ and $\psi = 0.8$, (c) $\phi = 0.8$ and $\psi = 0.8$, (d) $\phi = 0.3$ and $\psi = 0.8$

to distinguish between purely causal, noncausal or mixed processes as their autocovariance functions are identical. Fitting an autoregressive model by OLS allows however to estimate the sum of leads and lags, p.[3] Subsequently, the respective numbers of lags (r) and leads (s), such that $r + s = p$, can be estimated by an approximate maximum likelihood (hereafter AML) approach (Lanne and Saikkonen, 2011). The selected model is the one maximising the AML with respect to r, s and all parameters $\Omega = (\Phi, \Psi, \Theta)$, where $\Phi = (\phi_1, \ldots, \phi_r)$, $\Psi = (\psi_1, \ldots, \psi_s)$ and Θ is the errors distribution parameters, such as the scale or location for instance. The AML estimator is defined as follows,

$$\left(\hat{\Phi}, \hat{\Psi}, \hat{\Theta}\right) = argmax_{\Phi,\Psi,\Theta} \sum_{t=r+1}^{T-s} \ln\left[g\Big(\Phi(L)\Psi(L^{-1})y_t; \Theta\Big)\right],$$

where g denotes the *pdf* of the error term, satisfying the regularity con-

[3] A non-Gaussian MLE can give misleading results in a misspecified model (Gouriéroux and Jasiak, 2018).

ditions (Andrews, Davis, and Breidt, 2006). Lanne and Saikkonen (2011) show that the resulting (local) maximum likelihood estimator is consistent, asymptotically normal and that $(\hat{\Psi}, \hat{\Phi})$ and $\hat{\Theta}$ are asymptotically independent, for Student's t-distributed errors with finite moments. Since an analytic solution of the maximisation problem at hand is not directly available, numerical gradient-based procedures can be employed. Hecq et al. (2016) indicate that estimating MAR models is easier for more volatile series since the convergence of the estimator is empirically faster for distributions with fatter tails. They propose an alternative way to obtain the standard errors, a method implemented in the R package MARX (Hecq, Lieb, and Telg, 2017).

While with a unique lead and a positive coefficient ψ, explosive episodes increase at a fixed rate ψ^{-1} until a sudden crash, other specifications induce complex patterns not resembling the bubble pattern that this paper focuses on. This paper hence only considers $MAR(r,1)$ processes with a positive lead coefficient.

3 Predictions using closed-form expressions

When it comes to forecasting MAR processes, different approaches are available. Conditional expectations can be used to predict the next points, while alternatively, forecasting densities aims at visually analysing the probabilities of potential future paths. The latter can be employed to evaluate the probabilities of a turning point in an explosive episode. However, the anticipative aspect of MAR models complicates their use for predictions. Results are not as straightforward as they could be with purely backward-looking $ARMA$ models. While in some cases mean or density forecasts can be directly obtained from the assumed errors distribution, they sometimes need to be approximated. For this section, let us assume that the data generating process (hereafter *dgp*) is a stationary $MAR(r,1)$ process $\Phi(L)(1-\psi L^{-1})y_t = \varepsilon_t$, where $\psi > 0$, ε_t is i.i.d. non-Gaussian and $u_t = \Phi(L)y_t$ is the purely non-causal component of the process. Throughout the coming sections, the *dgp* is be assumed correctly identified to disregard estimation uncertainty.

Given the true *dgp*, the information sets $(y_1, \ldots, y_T, y^*_{T+1}, \ldots, y^*_{T+h})$ and $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-1}, u_T, u^*_{T+1}, \ldots, u^*_{T+h})$, where $v_t = \Phi(L)^{-1}\varepsilon_t$ and $u_t = (1-\psi L^{-1})^{-1}\varepsilon_t$, are equivalent. This allows to predict future values of y from predictions of the forward-looking component of ε, namely u.

The asterisk indicates unrealised values of the random variables. Most prediction methods hence aim attention at purely noncausal processes – here u_t, sufficient to predict the variable of interest.

3.1 Point predictions

Gouriéroux and Zakoïan (2017) derive the first two conditional moments of $MAR(0,1)$ processes,[4] here denoted as u_t, and show that for Cauchy processes, with $\psi > 0$, the expectation of u_{T+1} conditioned on the information set known at time T, $\mathcal{F}_T$, is

$$\mathbb{E}\big[u_{T+1}|\mathcal{F}_T\big] = u_T. \tag{3}$$

This result is puzzling since the conditional expectation of a noncausal process has a unit root even though the process is stationary. Fries (2018) expanded those results to any admissible parametrisation of the tail and asymmetry parameters of α-stable distributions and derives up to the fourth conditional moments. He also derives the limiting distribution of those four moments when the variable of interest diverges. He shows that during an explosive episode, the computation of those moments gets considerably simplified and are characteristic of a weighted Bernoulli distribution charging probability $\psi^{\alpha h}$ to the value $\psi^{-h}u_T$ and $(1 - \psi^{\alpha h})$ to value zero, for a tail parameter $0 < \alpha < 2$. Those results indicate that along a bubble, the process can only either keep on increasing with fixed rate or drop to zero. For Cauchy-distributed errors ($\alpha = 1$), the mean forecast during an explosive episode remains equal to Equation (3), yet for other α-stable distributions the conditional expectation may be drastically simplified. Hence, during an explosive episode, the point forecast of an $MAR(0,1)$ process is a weighted average of the crash and further increase (e.g. a random walk for Cauchy-distributed processes). Density forecasts may therefore be more informative.

3.2 Density predictions

The joint conditional predictive density (as named by Gouriéroux and Jasiak, 2016) or the causal transition distribution (as named by Gouriéroux and Zakoïan, 2017) of the h future values, $(u^*_{T+1}, \ldots, u^*_{T+h})$, given the information known at time T can be evaluated only conditioning on the last observed value u_T. This stems from the equivalence of information set of

[4]Note that the linear projection on the past does not correspond, in this context, to the conditional expectation (Gouriéroux and Jasiak, 2018).

the observed series and of its filtrations when the model is assumed correctly identified and the independence between the error components. While the interest is on predicting the future given present and past information, it is only possible, by the model definition, to derive the density of a point conditional on its future point. Bayes' Theorem is first used to get rid of the conditioning on the present point and a second time to condition on the last point of the forecast. Then, the theorem is applied repeatedly on the first term until the density of all points is conditional on future ones. The conditional *pdf* can thus be expressed as follows,

$$\begin{aligned}
&l(u^*_{T+1},\dots,u^*_{T+h}|u_T)\\
&= l(u_T,u^*_{T+1},\dots,u^*_{T+h-1}|u^*_{T+h}) \times \frac{l(u^*_{T+h})}{l(u_T)}\\
&= \left\{ l(u_T|u^*_{T+1},\dots,u^*_{T+h}) l(u^*_{T+1}|u^*_{T+2},\dots,u^*_{T+h}) \dots l(u^*_{T+h-1}|u^*_{T+h}) \right\}\\
&\qquad \times \frac{l(u^*_{T+h})}{l(u_T)},
\end{aligned}$$

where l denotes densities associated with the noncausal process u_t. Equation (2) states that $\varepsilon_t = u_t - \psi u_{t+1}$, hence, for all t, only u_{t+1} is necessary to derive u_t. Furthermore, given u_{t+1}, the conditional density of u_t (which is not known) is equivalent to the density of ε_t (which is known) evaluated at the point $u_t - \psi u_{t+1}$.[5] That is, for any assumed errors distribution g we have,

$$\begin{aligned}
&l(u^*_{T+1},\dots,u^*_{T+h}|u_T)\\
&= \left\{ l(u_T|u^*_{T+1}) l(u^*_{T+1}|u^*_{T+2}) \dots l(u^*_{T+h-1}|u^*_{T+h}) \right\} \times \frac{l(u^*_{T+h})}{l(u_T)}\\
&= g(u_T - \psi u^*_{T+1}) \dots g(u^*_{T+h-1} - \psi u^*_{T+h}) \times \frac{l(u^*_{T+h})}{l(u_T)}.
\end{aligned} \tag{4}$$

Problems may however arise when the stationary distribution of u_t is unknown. We know that $u_t = \psi u_{t+1} + \varepsilon_t = \sum_{i=0}^{\infty} \psi^i \varepsilon_{t+i}$, but the *pdf* of a linear combinations of errors may not admit closed-form expressions for some distributions.

[5] Since $u_\tau = \psi u_{\tau+1} + \varepsilon_\tau$ and because $u_{\tau+1}$ and ε_τ are independent for all $1 \le \tau \le T$, we have $f_{u_\tau|u_{\tau+1}}(x) = f_{\varepsilon_\tau+\psi u_{\tau+1}|u_{\tau+1}}(x) = f_{\varepsilon_\tau|u_{\tau+1}}(x - \psi u_{\tau+1}) = f_\varepsilon(x - \psi u_{\tau+1})$. For simplicity, the density distributions related to u_t (resp. ε_t) are just denoted by l (resp. g).

For instance, Gouriéroux and Zakoïan (2013) present closed-form solutions for the predictive conditional density of purely noncausal $MAR(0,1)$ processes with Cauchy-distributed errors. They show that the characteristic function of the infinite sum corresponds to that of a Cauchy with scale parameter $\frac{\gamma}{(1-\psi)}$, where γ is the scale of the distribution of the errors ε_t. Hence, in the $MAR(r,1)$ case with Cauchy errors, $u_t \sim Cauchy\Big(0, \frac{\gamma}{(1-\psi)}\Big)$. The predictive density of the purely noncausal process (u_t) can thus be computed as such,

$$
\begin{aligned}
&l(u^*_{T+1}, \ldots, u^*_{T+h}|u_T) \\
&= \frac{1}{(\pi\gamma)^h}\left(\frac{1}{1+\frac{(u_T-\psi u^*_{T+1})^2}{\gamma^2}} \cdots \frac{1}{1+\frac{(u^*_{T+h-1}-\psi u^*_{T+h})^2}{\gamma^2}}\right) \\
&\times \frac{\gamma^2+(1-\psi)^2u_T^2}{\gamma^2+(1-\psi)^2(u^*_{T+h})^2}.
\end{aligned}
$$

To illustrate how the predictive density evolves as the series departs from central values, Figure 2 shows one-step ahead forecasts for different levels corresponding to quantiles 0.55, 0.85 and 0.975 of a purely noncausal process with a lead coefficient of 0.8 and standard Cauchy-distributed errors. By using quantiles, explosive episodes can be compared between different distributions and parameters.[6] While the predictive distribution is uni-modal for median-level values, it splits and becomes bi-modal when the series departs from such values. As the series diverges, the bi-modality of the conditional distribution becomes more evident, where the two modes correspond to a drop to 0 and a continuous increase with rate (1/0.8); each event has probability 0.2 and 0.8 respectively. Those results corroborate what Fries (2018) shows for diverging Cauchy-distributed $MAR(0,1)$ series. Note that results are analogous for any parameters, with corresponding probabilities of a crash equal to $1-\psi$. Bi-modality in this paper will therefore designate the split of the conditional density and not the bi-modality sometimes observed in the estimation of the coefficients of MAR models (Hecq et al., 2016 and Bec, Bohn Nielsen, and Saïdi, 2019)

For $MAR(r,1)$ processes, one-step ahead density forecasts consists in shifting the predictive density of the purely non-causal component by the causal part of the process, namely $\phi_1 y_T + ... + \phi_r y_{T-r+1}$. For an h-step ahead forecast, with $h \geq 1$, the predictive density of y^*_{T+h} will depend on the joint density

[6] We thank an anonymous referee for the suggestion.

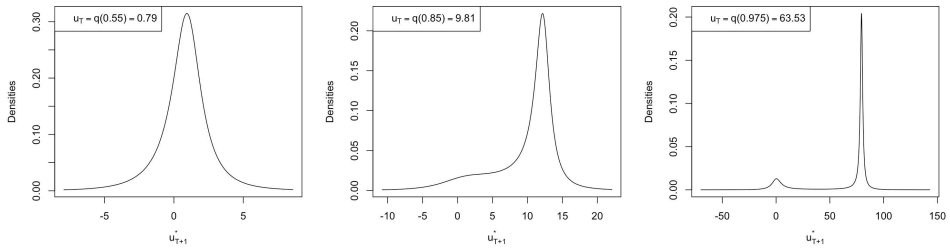


Figure 2: Evolution of the 1-step ahead predictive density as the level of the series increases for a Cauchy $MAR(0,1)$ with $\psi = 0.8$.

of $(u^*_{T+1}, \ldots, u^*_{T+h})$. One way of approaching this is to directly write the predictive density in terms of y^*_{T+k}, with $1 \leq k \leq h$, in the conditional joint density of $(u^*_{T+1}, \ldots, u^*_{T+h})$. For an $MAR(1,1)$ process with lag coefficient ϕ for instance, the conditional predictive density of h future y's could be obtained as follows,

$$
\begin{aligned}
l(y^*_{T+1}, \ldots, y^*_{T+h}|y_T) &= \frac{1}{(\pi\gamma)^h} \\
&\times \frac{1}{1 + \frac{(u_T - \psi(y^*_{T+1} - \phi y_T))^2}{\gamma^2}} \cdots \frac{1}{1 + \frac{(y^*_{T+h-1} - \phi y^*_{T+h-2}) - \psi(y^*_{T+h} - \phi y^*_{T+h-1}))^2}{\gamma^2}} \\
&\times \frac{\gamma^2 + (1-\psi)^2 u_T^2}{\gamma^2 + (1-\psi)^2 (y^*_{T+h} - \phi y^*_{T+h-1})^2}.
\end{aligned}
$$

Figure 3 shows the evolution of two-step ahead forecasts of a purely noncausal process with lead coefficient 0.8 and Cauchy-distributed errors as the variable increases. For high levels of the series, the split of the distribution is evident; at each step the series can either keep on increasing or drop to zero, where the latter corresponds to an absorbing state. An interpretation of the regions of the graphs with respect to potential future shapes of the series was given by Gouriéroux and Jasiak (2016).

Overall, density predictions yield a more complete forecast as they carry more information regarding potential future patterns of the series. They cannot be easily graphically displayed for forecast horizons larger than 2 as we investigate joint predictions, yet results can be used to compute the probabilities regarding future patterns. For instance, when the variable follows an explosive path, the probabilities of a crash can be computed from

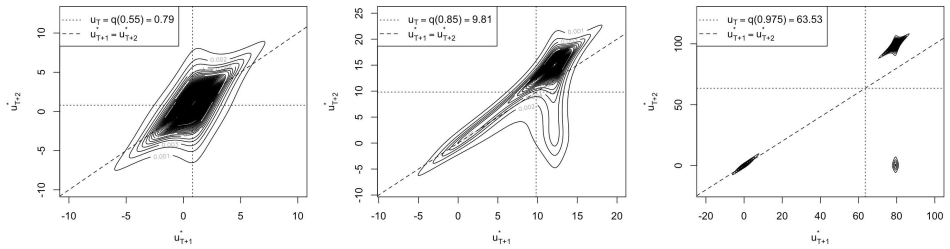


Figure 3: Evolution of the 2-step ahead joint predictive density as the level of the series increases for an $MAR(0,1)$ with $\psi = 0.8$ and Cauchy-distributed errors

the densities by choosing a threshold, such as the last observed value or its half for instance. Nonetheless, as indicated by Fries (2018) for α-stable distributions, explosive episodes seem to be memoryless and as the series diverges, the probabilities of a crash tend to the constant $1 - \psi^{\alpha h}$ for given a given horizon $h \geq 1$. Even though as $h \to \infty$ this probability tends to 1, it may not be very realistic when it comes to real life data. We might expect the probabilities of a crash in stock prices for instance to increase with the level of prices. Furthermore, the assumption of other fat-tail distributions (e.g. Student's t) generally leads to the absence of closed-form expressions for the conditional moments and densities. The next Section presents two approaches to approximate the conditional densities in such circumstances; the first one is based on simulations (Lanne, Luoto, and Saikkonen, 2012) and the second one uses sample counterparts (Gouriéroux and Jasiak, 2016).

4 Predictions using approximation methods

4.1 Predictions using simulations-based approximations

Lanne, Luoto, and Saikkonen (2012) base their methodology on the fact that the noncausal component of the errors, u, can be expressed as an infinite sum of future errors, which in the $MAR(r,1)$ case is as follows,

$$u_t = \Psi(L^{-1})^{-1}\varepsilon_t = \sum_{i=0}^{\infty} \psi^i \varepsilon_{t+i}.$$

Since stationarity is assumed, and because in applications ψ rarely (and barely) exceeds 0.9, the sequence (ψ^i) converges rapidly to zero. Hence, they assumed that there exists an integer M large enough so that any future

point of the noncausal component of the errors can be approximated as the following finite sum,

$$u^*_{T+h} \approx \sum_{i=0}^{M-h} \psi^i \varepsilon^*_{T+h+i}, \tag{5}$$

for any $h \geq 1$.

As explained before, any point forecast y^*_{T+h} of an MAR(r,1) process depends on the sequence forecast $(u^*_{T+1}, \ldots, u^*_{T+h})$. Thus, forecasting any future point y^*_{T+h} or the path $(y^*_{T+1}, \ldots, y^*_{T+h})$, with $h \geq 1$, requires forecasting the sequence of M future errors $(\varepsilon^*_{T+1}, \ldots, \varepsilon^*_{T+M})$ which we will denote as ε^*_+. Instead of deriving an M-dimensional conditional joint density function (Lanne, Luoto, and Saikkonen (2012) use $M = 50$) , they propose a way to obtain conditional point and cumulative density forecasts. While the estimation approach they propose requires finite moments for the errors distribution, this restriction is not necessary for their forecasting method (Lanne and Saikkonen, 2011).

Using the companion form of an $MAR(r,1)$ model, y^*_{T+h} can, by recursion, be expressed as the sum of a known component and the h future values of u_t, where the latter, based on Equation (5), can be approximated as a linear combination of M future errors

$$\begin{aligned} y^*_{T+h} &= \iota' \mathbf{\Phi}^h \mathbf{y}_T + \sum_{i=0}^{h-1} \iota' \mathbf{\Phi}^i \iota u^*_{T+h-i} \\ &\approx \iota' \mathbf{\Phi}^h \mathbf{y}_T + \sum_{i=0}^{h-1} \iota' \mathbf{\Phi}^i \iota \sum_{j=0}^{M-h+i} \psi^j \varepsilon^*_{T+h-i+j}, \end{aligned} \tag{6}$$

where

$$\mathbf{y}_T = \begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-r+1} \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \ldots & \phi_r \\ 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{bmatrix} (r \times r) \quad \text{and} \quad \iota = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (r \times 1).$$

Let $g(\varepsilon^*_+|u_T)$ be the conditional joint distribution of the M future errors, which, using Bayes' Theorem can be expressed as follows,

$$g(\varepsilon^*_+|u_T) = \frac{l(u_T|\varepsilon^*_+)}{l(u_T)} g(\varepsilon^*_+).$$

Thus, for any function q,

$$\begin{aligned}\mathbb{E}\Big[q(\varepsilon_+^*)\big|u_T\Big] &= \int q(\varepsilon_+^*)g(\varepsilon_+^*|u_T)d\varepsilon_+^* \\ &= \frac{1}{l(u_T)}\int q(\varepsilon_+^*)l(u_T|\varepsilon_+^*)g(\varepsilon_+^*)d\varepsilon_+^* \\ &= \frac{\mathbb{E}_{\varepsilon_+^*}\Big[q(\varepsilon_+^*)l(u_T|\varepsilon_+^*)\Big]}{l(u_T)}.\end{aligned} \tag{7}$$

Similarly as before, $l(u_T|\varepsilon_+^*)$ can be obtained from the errors distribution g. Yet, since it is conditional on ε_+^* instead of u_{T+1}^*, we can only obtain an approximation. Using this approximation and the Iterated Expectation theorem, the marginal distribution of u_T can be approximated as follows,

$$l(u_T) = \mathbb{E}_{\varepsilon_+^*}\big[l(u_T|\varepsilon_+^*)\big] \approx \mathbb{E}_{\varepsilon_+^*}\left[g\left(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\right)\right].$$

Overall, by plugging the aforementioned approximation in (7), we obtain

$$\mathbb{E}\Big[q(\varepsilon_+^*)\big|u_T\Big] \approx \frac{\mathbb{E}_{\varepsilon_+^*}\left[q(\varepsilon_+^*)g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\Big)\right]}{\mathbb{E}_{\varepsilon_+^*}\left[g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\Big)\right]}.$$

Let $\varepsilon_+^{*(j)} = \Big(\varepsilon_{T+1}^{*(j)}, \ldots, \varepsilon_{T+M}^{*(j)}\Big)$, with $1 \leq j \leq N$, be the j-th simulated series of M independent errors, randomly drawn from the assumed distribution of the process. Assuming that the number of simulations N is large enough, the conditional expectation of interest can be approximated as follows,

$$\mathbb{E}\Big[q(\varepsilon_+^*)\big|u_T\Big] \approx \frac{N^{-1}\sum_{j=1}^{N} q\Big(\varepsilon_+^{*(j)}\Big)g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^{*(j)}\Big)}{N^{-1}\sum_{j=1}^{N} g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^{*(j)}\Big)}. \tag{8}$$

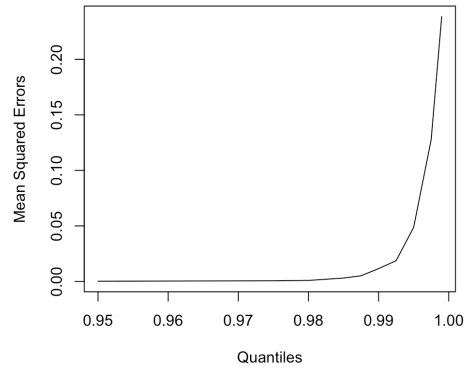
Based on Equation (6), for any $MAR(r,1)$ process and for any forecast horizon $h \geq 1$, choosing $q(\varepsilon_+^*) = \mathbf{1}\big(\sum_{i=0}^{h-1}\iota'\mathbf{\Phi}^i\iota\sum_{j=0}^{M-h+i}\psi^j\varepsilon_{T+h-i+j}^* \leq x - \iota'\mathbf{\Phi}^h\mathbf{y}_T\big)$ in (8) will provide an approximation of $\mathbb{P}\big(y_{T+h}^* \leq x|u_T\big)$. By computing its value for all possible x we can obtain the whole conditional

cdf of y^*_{T+h}.

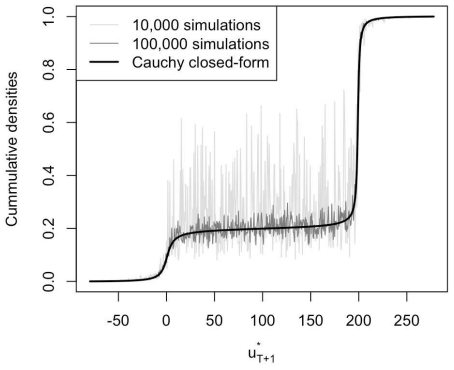
$$\begin{aligned}\mathbb{P}\Big(y^*_{T+h} \leq x \big| u_T\Big) &= \mathbb{E}\Big[\mathbf{1}\big(y^*_{T+h} \leq x\big)\big| u_T\Big] \\ &\approx \mathbb{E}\left[\mathbf{1}\left(\iota'\mathbf{\Phi}^h\mathbf{y}_T + \sum_{i=0}^{h-1}\iota'\mathbf{\Phi}^i\iota\sum_{j=0}^{M-h+i}\psi^j\varepsilon^*_{T+h-i+j} \leq x\right)\Big| u_T\right].\end{aligned}$$

Let us consider again an $MAR(0,1)$ process with a lead coefficient of 0.8 and Cauchy-distributed errors. The complete predictive *cdf* is approximated using $M = 100$ and the 10,000 simulations suggested by Lanne, Luoto, and Saikkonen (2012) at each iteration. The Mean Squared Errors (henceforth MSE) between the estimated and the theoretical *cdf*'s for increasing quantile (between the $Q(0.95)$ and $Q(0.999)$) of the MAR process are presented on graph (a) in Figure 4. The MSEs increase with the level of the series (from 0.0002 to 0.2384) and for illustration, graph (b) in Figure 4 compares the *cdf*'s obtained with 10,000 and 100,000 simulations with the theoretical *cdf* for quantile 0.99. The discrepancy between the estimated and theoretical *cdf*'s significantly decrease with a larger number of simulations and results converge towards the theoretical distribution. Furthermore, it is important to note that the bi-modality of the conditional distribution of an explosive episode is captured by this approach.

Figure 5 depicts the empirical distribution of 1,000 iterations of the same inquiry with different number of simulations. Since this paper focuses on the investigation of turning point of explosive episodes, each iteration consists in computing the probabilities of a decrease of at least 25% when the last observed value is equal to the quantile 0.995, namely a value of around 318 for such process. The theoretical probability is equal to 0.2 and as the number of simulations increases, results converge to this value. For lower number of simulations the same inquiry repeated twice may give completely opposite results. This stems from the fact that we investigate explosive episodes. Recall that bubbles are triggered by a future extreme value in the error terms and if no simulated paths among all simulations can trigger such increase, then the probabilities may be significantly misleading. Conclusions are similar for different lead coefficients (results available upon request), the simulations-based probabilities are a good approximation of theoretical Cauchy-derived probabilities ($(1-\psi)$ during explosive episodes), when the number of simulations is coherently chosen w.r.t. the level of the series.



(a) Evolution of the MSE of the *cdf* estimations with 10,000 simulations for increasing quantiles.



(b) Comparison of estimated *cdf*'s for Q(0.99) using 10,000 and 100,000 simulations with theoretical *cdf*.

Figure 4: Sensitivity of estimations to the number of simulations for an MAR(0,1) with $\psi = 0.8$ and Cauchy distributed errors.

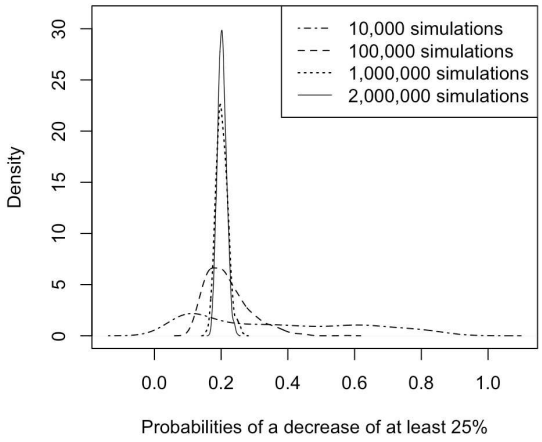


Figure 5: Empirical distributions of 1,000 repeated forecasts using different numbers of simulations for an MAR(0,1) process with $\psi = 0.8$ and Cauchy distributed errors evaluated at quantile 0.995.

Analogously, results also converge for Student's t-distributed processes – we here investigate $t(2)$- and $t(3)$-distributed $MAR(0,1)$ series. No

theoretical results are available but Figure 6 indicate that results converge to a unique distribution as the number of simulations is increased in the estimation. Values corresponding to similar quantiles significantly vary with the distributions. While for a Cauchy ($t(1)$) distributed $MAR(0,1)$ process with $\psi = 0.8$, Q(0.995) corresponds to a value of 318.28, it corresponds to 17.35 and 8.75 for $t(2)$ and $t(3)$ respectively.[7] Hence, since the rate of increase remains the same (1/0.8), the modes of the conditional distribution are closer and the bi-modality is less evident for similar quantiles when the degrees of freedom of the distribution are larger. Hence, given a quantile, probabilities of events (e.g. drop of at least 25% or of at least 50%) will differ the most when the quantile corresponds to lower values. Furthermore, for analogous quantiles, approximations are less sensitive to the number of simulations as the degrees of freedom of the distribution increases. This is explained by the fact that the values needed to be drawn from the distribution to keep on following the current explosion rate (1/0.8) do not correspond to the same quantiles. Namely, $(1/0.8)Q(0.995)$ does not correspond to the same quantile depending on the distribution. For $t(1)$ the rate of increase would lead to quantile 0.999 while for $t(3)$ it would lead to quantile 0.986. Hence, when reaching the same quantile, it is more likely that the values corresponding to the natural rate of increase are simulated when the degrees of freedom of the distribution are larger.

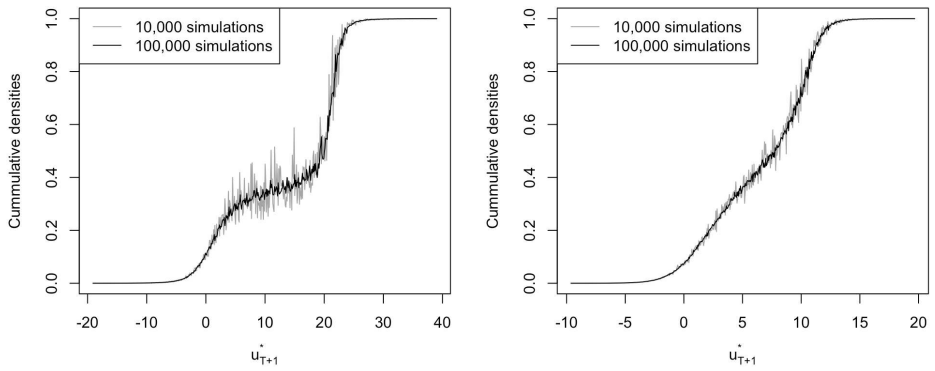


Figure 6: Estimated *cdf*'s evaluated at Q(0.995) using 10,000 and 100,000 simulations for MAR(0,1) processes with $\psi = 0.8$ with $t(2)$ (left) and $t(3)$ (right) distributed errors.

[7]Note that quantiles for t-distributed processes were empirically estimated.

To compare results with Cauchy-distributed processes, Table 1 displays the probabilities of a decrease of at least 25% once the quantile 0.995 is attained for the three distributions and for three distinct lead coefficients. Theoretical probabilities from the same quantile are also reported for Cauchy ($t(1)$) distributed errors. Simulations-based probabilities differ by a maximum of 0.2% from theoretical results in the $t(1)$ case. Furthermore, recall that the larger the degrees of freedom, the less noisy are estimations for a given number of simulations. That is, we can expect the probabilities for the $t(2)$ and $t(3)$ cases to differ from theoretical probabilities by at most 0.2% and further increasing the number of simulations would lead to more precise results. Given the same quantile, the probabilities of a turning point significantly increases with the degrees of freedom of the distribution and with lower lead coefficients. For the investigated quantile, a process with $t(3)$-distributed errors and a lead coefficient of 0.2 only has a probability of 4% to keep on increasing as opposed to 20% for Cauchy-distributed processes. Furthermore, simulations indicate that as the series diverges, the probabilities of a crash (impact of the choice of threshold becomes negligible as the modes of the distribution departs from one another) tend to a constant for all models. For a lead coefficient of 0.8 for instance, the probabilities of a downturn tends to 0.2, 0.36 and 0.48 as the bubble increases for $t(1)$, $t(2)$ and $t(3)$ respectively.

Table 1: Probabilities of a crash of at least 25% when quantile 0.995 is attained

Lead coefficient	Theoretical	Simulations-based		
ψ	Cauchy/$t(1)$	$t(1)$	$t(2)$	$t(3)$
0.2	.794	.793	.937	.960
0.5	.497	.499	.718	.792
0.8	.201	.203	.358	.435

Reported probabilities for the simulations-based approach are the average over 1,000 forecasts using 1,000,000 simulations.

Overall, this approach seems to be a good approximation of theoretical results. With a sufficient number of simulations, the probabilities obtained with Cauchy-distributed errors converge to theoretical probabilities. While for $t(2)$ and $t(3)$ results cannot be compared to a benchmark, estimated distributions also converge to a unique distribution which we can expect to be the theoretical one. Nonetheless, obtained probabilities do not depend on past behaviours and, as was indicated in Section 3 for theoretical results,

tend to a constant during explosive episodes. That is, past some point, the probabilities of a crash will remain constant.

4.2 Predictions using sample-based approximations

This section is based on the approach proposed by Gouriéroux and Jasiak (2016). They derive a sample-based estimator of the ratio of the predictive densities in Equation (4), which does not always admit closed-form results. Based on past values of the series, this method can also be applied to any non-Gaussian distribution. Whether or not the marginal distributions of u_t and u^*_{T+h} admit closed-form, they can be expressed as follows,

$$l(u_\tau) = \mathbb{E}_{u_{\tau+1}}\big[l(u_\tau|u_{\tau+1})\big],$$

with $\tau = \{T, T+h\}$. Once again the noncausal relationship described in Equation (2) is used to evaluate the conditional distribution of $l(u_\tau|u_{\tau+1})$ with the distribution of the errors, $g(u_\tau - \psi u_{\tau+1})$. While Lanne, Luoto, and Saikkonen (2012) employed simulations to approximate expected values, Gouriéroux and Jasiak (2016) use sample-based counterparts. The expected value here is approximated by the average obtained using all points from the sample for the conditioning variable,

$$l(u_\tau) = \mathbb{E}_{u_{\tau+1}}\big[g(u_\tau - \psi u_{\tau+1})\big] \approx \frac{1}{T}\sum_{i=1}^{T}\Big\{g(u_\tau - \psi u_i)\Big\}. \tag{9}$$

Hence, the predictive density for the $MAR(0,1)$ process u_t can be approximated by plugging the sample counterparts (9) in (4),

$$\begin{aligned} &l(u^*_{T+1}, \ldots, u^*_{T+h}|u_T) \\ &\approx g(u_T - \psi u^*_{T+1}) \ldots g(u^*_{T+h-1} - \psi u^*_{T+h}) \frac{\sum_{i=1}^{T} g(u^*_{T+h} - \psi u_i)}{\sum_{i=1}^{T} g(u_T - \psi u_i)}. \end{aligned} \tag{10}$$

For centred and symmetrical uni-modal distributions, such as the Cauchy and the Student's t that are employed in this analysis, the probability density function is maximised at zero. That is, the density g, as it is evaluated in Equation (10), is maximised at the points where $u_\tau - \psi u_i = 0$ and tends to zero as the difference widens. Since at time T all observations up to u_T are used in the estimation, the ratio of Equation (10) only varies

as a function of u^*_{T+h} and will be maximised for paths that were already undertaken. Furthermore, when u_T diverges from all past values in the sample, the numerator tends to zero, meaning that approximations errors will be amplified during explosive episodes. That is, we can expect this approximation method to put more weight on forecast points corresponding to already undertaken paths and that this tendency will be more pronounced during bubbles.

Let us again consider an $MAR(0,1)$ process with a lead coefficient of 0.8 and standard Cauchy-distributed errors. For median levels of the series, results are similar between closed-form and sample-based predictions regardless of past behaviours. However, as the series departs from central values, discrepancies emerge and are path-dependent. To illustrate this, Figure 7 shows one-step ahead density forecasts performed at time *T=200* of two different $MAR(0,1)$ trajectories, both ending at the same point corresponding to quantile 0.975 (equivalent to a level of 63.53 for such model). Series 1 (left) only has smaller explosive episodes before the one at which predictions are performed while series 2 already lived a more considerable bubble before. The one-step ahead density predictions are estimated and compared to closed-form Cauchy results in the bottom graphs. The estimator captures the split of the density of the explosive episode but the densities are significantly different. Compared to the theoretical density (solid line), the left estimated density overestimates the probabilities related to a crash since all past points are lower, while the right one overestimates potential increases as the series already underwent larger explosive episodes. Predictions with this approach are therefore case-specific and can be characterised by a learning mechanism based on past behaviours. The probabilities of events can be empirically derived from the obtained predictive densities. The probabilities of a decrease of at least 25% are theoretically 20.5% for such process but are respectively equal to 55.7% and 26.3% for series 1 and 2, for which the discrepancy is explained by the aforementioned learning mechanism. The choice of event and thus threshold used to calculate the probabilities may have a considerable impact on the results. Theoretically, for such process and quantile, the probabilities of a drop of 75% are only 3.2% lower than for a drop of 25%. This indicates that the arbitrary definition of a crash (e.g. a drop of 50% or of 25%) does not significantly affect the resulting probabilities. However, for series 1 (resp. 2), the probabilities of a drop of at least 75% are 11% (resp. 5%) lower. That is, the learning mechanism can induce substantial probabilities for scenarios in between the crash and the further increase. Hence, caution

is needed when building probability-based investment strategies for instance.

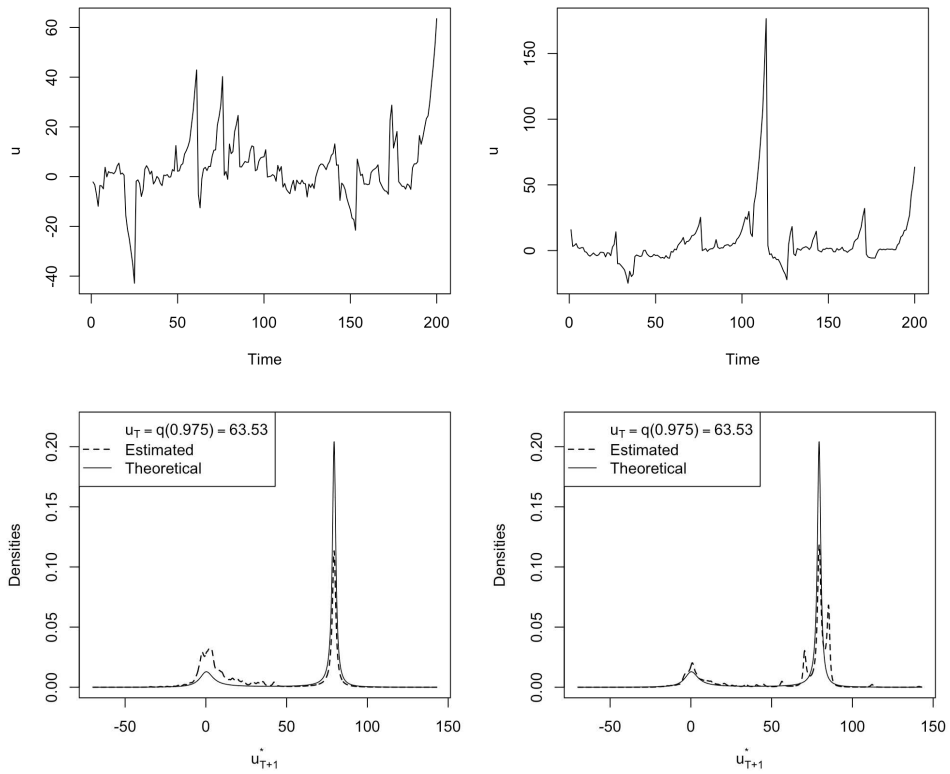


Figure 7: Comparison between estimated and theoretical 1-step ahead predictive densities for Cauchy $MAR(0,1)$ with $\psi = 0.8$ evaluated at quantile 0.975 for 2 distinct trajectories.

In the intent of generalising results, we simulated sets of 1,000 different MAR(0,1) (with Cauchy-distributed errors) trajectories ending at quantiles 0.99 or 0.995, with lead coefficients 0.2, 0.5 or 0.8 and sample size 100, 200, 500 or 1000. The probabilities of a decrease of at least 25% were computed for all settings and all trajectories. Figure 8 reports the distribution of results depending on lead coefficient, sample size (500 on the left column and 1,000 on the right) and quantile investigated (0.99 and 0.995). The probabilities obtained from different trajectories are more volatile for large lead coefficients, large sample size and low quantiles. With such sample sizes, the empirical quantile corresponding to the last point when the lead coefficient is large is usually lower than the theoretical quantile

due to long lasting bubbles inducing substantial discrepancies between trajectories. This also implies that a larger proportion of points are of higher magnitude, which, as explained above may significantly alter probabilities, hence inducing volatility in the results. The same goes for lower the quantiles, a larger proportion of points of higher magnitude amplifies divergence between probabilities of two distinct trajectories. Increasing the sample size increases the occurrence of extreme episodes which also considerably affect probabilities. Indeed, as we have seen in Figure 7 one previous extreme episode is sufficient to significantly decrease the probabilities of a crash. Furthermore, we can see that the larger the coefficient, the more the sample-based approach tends to overestimates probabilities, compared to theoretical ones (represented by the dotted line). Note however, that the maximum probabilities obtained correspond to the main mode of the distributions and that the divergence of results (resp. the change of quantile) only happen in or (resp. affect) the left tail. Compared to theoretical probabilities, this approach tend to overestimates the probabilities of a crash but results seem to be upper-bounded, and this upper bound corresponds to the most recurrent obtained probability over 1,000 trajectories. However, probabilities of a crash can also be lower than theoretical probabilities, that is, the learning mechanism can indicate that based on past behaviours, the probabilities of turning point are lower than the underlying distribution would suggest.

Table 2 summarises the aforementioned results for four different samples sizes and the three distributions investigated in this paper. Obtained probabilities are compared to theoretical probabilities (for $t(2)$ and $t(3)$ results obtained in Table 1 are assumed to be theoretical ones). For each model setting, the first quantile and the mode of the distribution of the probabilities from the 1,000 trajectories are reported. The first quantile indicates the heaviness of the left tail while the mode indicates the upper-bound and by definition the most recurrent probability. Note that the tendency of the sample-based approach to overestimate the probabilities of a crash is lower for larger degrees of freedom in the errors distribution. Conclusions drawn above are analogous for all distribution, namely that an increase in the sample size and in the lead coefficient leads to significantly more volatile results. Nevertheless, due to the heavy dependence on past points and the case-specificity of this approach, it is rather challenging to demonstrate theoretical guarantees or convergence of this approximation method.

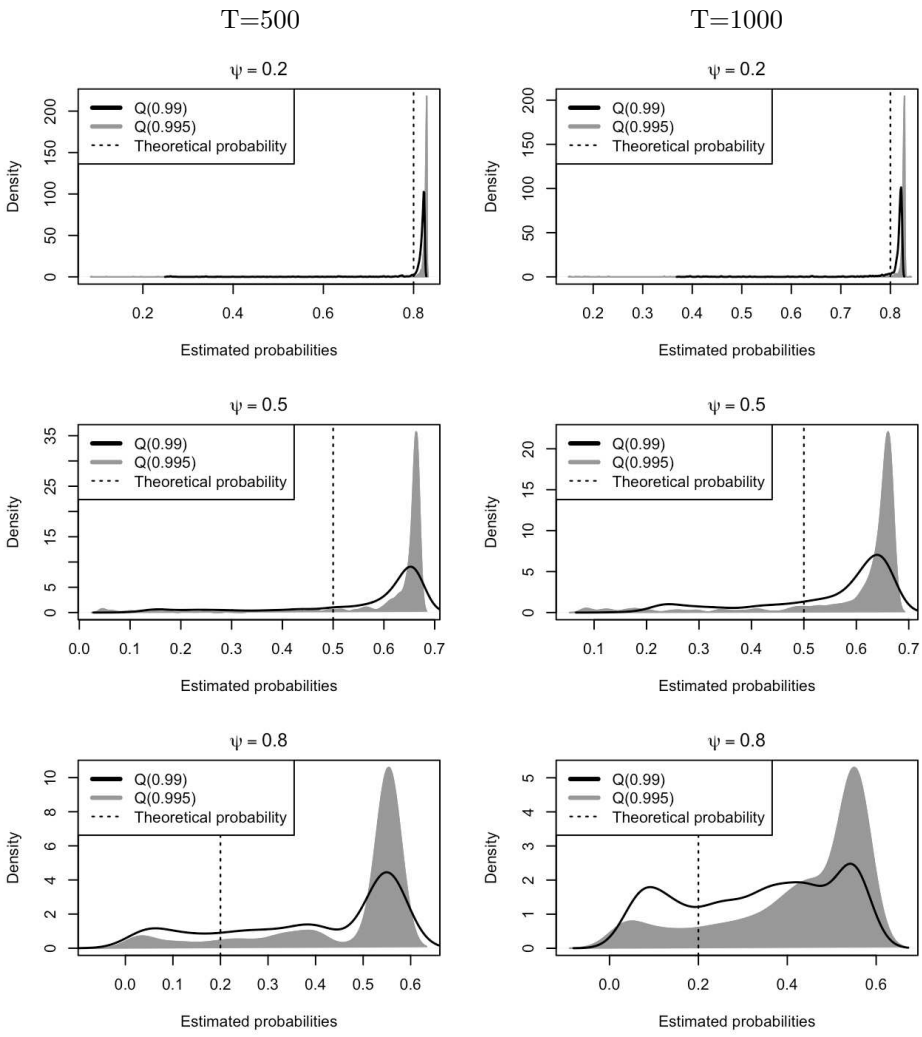


Figure 8: Distributions of estimated probabilities of a crash of at least 25% for 1,000 different trajectories evaluated at two different quantiles. The dotted lines represent theoretical Cauchy-derived probabilities. The lead coefficient varies (by row) and so does the sample size (by column).

The focus of this paper is on one-step ahead forecasts yet farther predictions are possible, though computationally demanding. Gouriéroux and Jasiak (2016) propose a method to tackle this issue by elaborating a Sampling Importance Resampling (SIR) algorithm. The algorithm aims at recovering a predictive density based on simulations from a misspecified instrumental

Table 2: Sample-based probabilities of a crash of at least 25% evaluated at Q(0.995) for 1,000 trajectories for each model

ψ	Sample size	$t(1)$			$t(2)$			$t(3)$		
		Theor.	1st Q.	Mode	Theor.*	1st Q.	Mode	Theor.*	1st Q.	Mode
0.2	100	.794	.828	.828	.937	.941	.941	.960	.961	.962
	200		.828	.828		.941	.941		.961	.961
	500		.825	.828		.941	.941		.961	.961
	1000		.824	.828		.940	.941		.961	.961
0.5	100	.497	.664	.665	.718	.772	.776	.792	.808	.819
	200		.657	.665		.756	.776		.802	.818
	500		.629	.665		.738	.775		.795	.819
	1000		.611	.665		.714	.775		.789	.815
0.8	100	.201	.555	.556	.358	.544	.597	.435	.428	.606
	200		.553	.556		.458	.603		.391	.590
	500		.404	.556		.259	.607		.369	.491
	1000		.343	.556		.290	.605		.379	.416

Theor. corresponds to theoretical probabilities (Theor.* correspond to the probabilities that were derived via simulations in the previous Section reported in Table 1).
Are also reported the 1st quantile and the mode of the distribution of the 1,000 probabilities of each scenario.

model from which it is easier to simulate. They suggest using a Gaussian AR model of order s (here an $AR(1)$) to simulate the process u_t. This approach recovers the intended densities for median levels of the series but fails to recover both the parts corresponding to the crash and to the increase during explosive episodes. The failure of the algorithm for high levels of the series stems from the intention to recover a bi-modal distribution from a uni-modal distribution. If the variance of the uni-modal instrumental distribution is not large enough to cover both modes of the sample-based density, the algorithm will not be able to recover the whole conditional distribution. The shape of the Normal distribution significantly depends on past behaviours of the series since the variance is estimated as the variance of the residuals of the MAR model. Hence, for more volatile series, the variance of the instrumental Normal distribution will be larger, yet, as the variable increases and the two modes diverge, there will always be a point from which the SIR algorithm does not succeed in recovering the density anymore.

Gouriéroux et al. (2018) find that the quality of forecasts diminishes when

the series follows an explosive episode. Indeed, approximations errors amplify with the level of the series, and there is a point from which the SIR algorithm does not recover the whole density anymore. Yet, we find that the sample-based estimator captures the split of the conditional density as the series departs from central values and comprises both the crash and increase parts of the predictive density. Furthermore, it yields time varying probabilities based on its learning mechanism. While sample-based predictive densities based on Student's t-distributions cannot be compared to closed-form predictions, results corroborate the conclusions drawn with Cauchy. Thinner tails in the errors distribution lead to higher probabilities of crash for given quantiles of an MAR process. A limitation is that when closed-form results are not available, we cannot disentangle how much of the derived probabilities are induced by the underlying distribution and how much by past behaviours. To tackle this, the probabilities estimated with the simulations-based approach of Lanne, Luoto, and Saikkonen (2012) can be used as benchmark as they seem to be good approximation of theoretical results. Such data-driven approach alleviates the issue of constant probabilities that theory or the simulations-based method suggest during explosive episodes. Yet, this is at the costs of heavy computations (increasing with the forecast horizon) and of lack of theoretical guarantees.

5 Empirical Analysis

We now empirically analyse the two approaches presented in Section 4. Karapanagiotidis (2014) and (Lof and Nyberg, 2017) find evidence that non-causal models generally provide better fits for commodity prices series. We hence forecast the bubble pattern in commodity prices and in particular in the monthly Global price of Nickel. The series is obtained from the International Monetary Fund and spans the period from January 1980 to June 2017. There seems to be a positive trend in the data but making the series stationary is far from obvious. Indeed, usual unit root tests do not perform well for this type of variable with very large spikes. For instance ADF tests would reject the null of a unit root against both a mean and a trend reverting alternative. A conclusion that does not seem satisfactory from the graphs of the data. It might also well be that the series is stationary around a shift in mean. Hencic and Gouriéroux (2015) use a cubic deterministic trend for isolating the bubble in the Bitcoin. In order to preserve the bubble features of the data and to obtain a stationary series

with locally explosive episodes (that would disappear by taking the returns) we have instead considered the Hodrick-Prescott filtering approach. The detrended series is reported in Figure 9. We are of course aware that this first step might alter the dynamics of the series, probably in the same manner that a X-11 seasonal filter modifies MAR models (see Hecq, Telg, and Lieb, 2017). We leave this important issue for further research. We first estimate an autoregressive model by OLS on the whole HP-detrended Nickel price series. Information criteria (AIC,BIC and HQ) all pick up a pseudo lag length of $p = 2$. The three possible $MAR(r,s)$ specifications are consequently an $MAR(2,0)$, an $MAR(1,1)$ or a $MAR(0,2)$. Using the MARX package of Hecq, Lieb, and Telg (2017) an $MAR(1,1)$ with a t-distribution with a degree of freedom of 1.32 and a scale parameter of 347.96 is favoured. The value of the causal and the noncausal parameters are respectively 0.60 and 0.74. We are consequently in the situation in which the predictive density does not admit closed-form expressions (although not very far from the Cauchy) but the sample- and simulations-based approaches can be used.

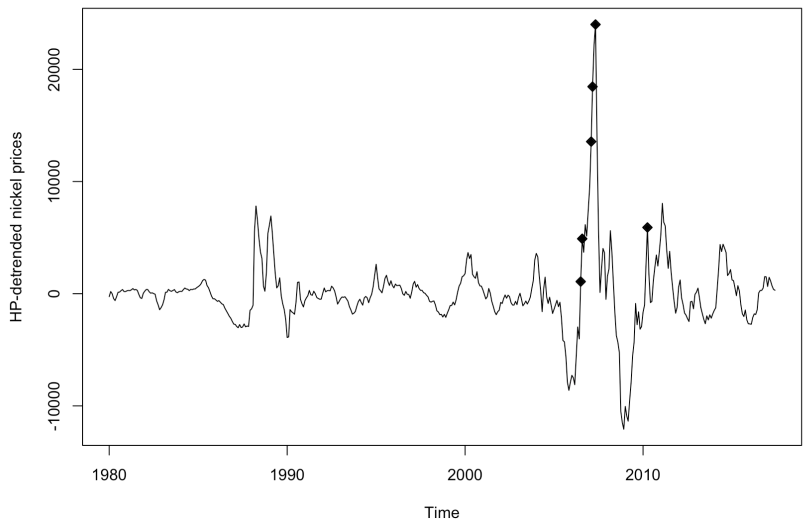


Figure 9: HP-detrended monthly Nickel prices series. The diamonds represent points from which one-step ahead density forecasts are performed in this analysis.

We aim attention at the main explosive episode, which crashed in June 2007. To investigate the evolution of predicted probabilities along the bubble

with settings as close as possible to the assumptions made throughout this paper, we assume the model is correctly specified (parameters estimated over the whole sample) at each point of interest. The points at which we perform predictions are represented by diamonds on the trajectory in Figure 9. We investigate five points along the main explosive episode and one after, to capture the effects of the inclusion of the crash in the predictions. Each point is assigned an index between 1 and 6 indicating their order of arrival. At each point, we compute the sample-based predictive density and compute various probabilities of events (four different magnitudes of crash) derived from both the sample- and simulations-based approaches. Since simulations-based estimations are good approximations (with a large enough number of simulations) of theoretical results, we consider them as theoretical benchmark to which sample-based probabilities are compared.

Results are reported in Table 3. The quantiles corresponding to each of the six points were evaluated using simulations, based on the estimated model, and are presented in the second column. The whole sample up to the points of interest were used in the sample-based approach. For the simulations-based method, given the degrees of freedom estimated for the errors and the quantiles to be investigated, 5,000,000 simulations were employed at each iteration. We investigate the probabilities of a decrease up to 60%. We do not consider larger drops since with a lag coefficient of 0.60, the left mode of the conditional distribution will be located at 60% of the last observed value and, as depicted in the last two columns, the probabilities of larger decrease will quickly decay to zero. Hence, let us now disregard the last two columns for the analysis of the results.

Points 1 to 5 represent the evolution from the outset to the peak of the bubble. During this episode the series departs from slightly above median values (Q(0.518)) to reach quantile 0.988. While the sample-based approach always overestimates theoretical probabilities, discrepancies between the two approaches widen as the series increases. More specifically, when reaching point 3, the series has now exceeded all past values and the discrepancy between probabilities of a drop with the two methods expands by 5.1% (a difference of 21.1% at point 2 and of 26.2% at point 3) and remains as such until the crash. The difference between them represents how much of the sample-based probabilities are induced by the learning mechanism of this approach. This suggests that once past values are exceeded, the uncertainty added to what the underlying distribution would suggest remains constant. The probabilities of a drop are strictly increasing with both approaches,

however, as mentioned in Section 4 the choice of threshold may impact probabilities. While this usually concerns the sample-based approach, this is also the case for low quantiles for the simulations-based method as the two modes of the conditional distribution are not sufficiently far to neglect the impact of the choice of threshold. The probabilities of a decrease rose by 1.3% with the sample-based approach between points 4 and 5 while probabilities of a crash of at least 25% declined by 0.1%, due to larger probabilities of drops of lower magnitudes evaluated at point 5.

Point 6 corresponds to a quantile slightly higher than point 2, hence, as expected, theoretical probabilities of a decrease are slightly higher as well. However, as the main explosive episode is now included in the sample-based approximations, the learning mechanism suggests less risk of a crash as the series already departed from such values. This leads to probabilities 12.2% lower than at point 2 for probabilities of a decrease and 9.8% lower for a crash of at least 25%. Furthermore, the discrepancy between the sample- and simulations-based approaches is reduced by more than 11% for both inquiries.

Table 3: One-step ahead probabilities of events for detrended monthly Nickel prices at six different point in time

Point	Quantile	$< y_T$		$< 75\%y_T$		$< 60\%y_T$		$< 40\%y_T$	
		samp.	sims.	samp.	sims.	samp.	sims.	samp.	sims.
1	.518	.291	.147	.227	.116	.185	.097	.131	.075
2	.902	.484	.273	.385	.183	.221	.105	.041	.033
3	.975	.587	.325	.518	.272	.273	.130	.008	.010
4	.984	.591	.330	.539	.289	.294	.139	.004	.006
5	.988	.604	.342	.538	.143	.291	.143	.002	.004
6	.923	.362	.289	.287	.199	.161	.107	.028	.027

The quantiles corresponding to each point were evaluated with simulations based on the estimated model.
samp. (resp. sims.) represents sample-based (resp. simulations-based) probabilities.
The probabilities are estimated with the following MAR(1,1) model: $\phi = 0.60$, $\psi = 0.74$ and $t(1.32,347.96)$ distributed errors.
For the simulations-based approach the following settings were employed: $M = 100$ and $N = 5{,}000{,}000$. Results with this approach are assumed to be theoretical ones.

5.1 Example of investment strategy

Using a monthly series to illustrate investment strategies may not be adequate but this is an example as to how the obtained probabilities could be employed and could be straightforwardly extended to higher frequencies. In addition, while the choice of the de-trending method may alter interpretation, it keeps the locally explosive characteristic of the series and is therefore a good illustrative example. A rather risk averse investor would probably sell once the series reaches point 3, that is when the learning mechanism indicates an increase of the probability of a turning point of 10.3% and more particularly an increase in the probabilities of a drop of at least 25% of 13.3%. Furthermore, as mentioned before, once exceeding all past values, the uncertainty carried by the sample-based approach is translated into a larger discrepancy between the sample- and simulations-based approaches (which then remains rather constant). If this investor did not sell at point 3, they would most likely sell at point 4. At that point, while the learning mechanism does not inflate the probabilities of a turning point (constant discrepancy between the two approaches), it still inflates the probabilities of a drop of at least 25%. This means that going from point 3 to 4, with the same excess probabilities of a turning point, if the series actually drops, higher probabilities are that it will drop by at least 25% when reaching point 4. A more risk seeking investor, if they have not sold yet at point 4 may be misled at point 5, indeed, excess probabilities of a turning point would still be the same that at point 4 while indicating that if the series drops it would less likely drop by more than 25% compared to the previous points. Moreover, any investment strategy using the learning mechanism of the sample-based approach would most probably lead to not selling at point 6 even though the small-scaled bubble crashed the subsequent point. Overall, the extent to which the excess probabilities (discrepancy between the sample- and simulations-based approaches) are used in an investment strategy depends on the beliefs regarding the series. The definition of a crash (namely just a turning point or a drop of at least 25% for instance) depends on the risk aversion of the investor. Further research should be done to investigate more thoroughly structured investment strategies and their performance over larger and diversified data sets.

To put it in a nutshell, the use of both approaches when the distribution of the errors does not allow for closed-form expressions can help disentangle

how much probabilities in the sample-based approach are induced by past behaviours. Indeed, even if probabilities of a turning point were continuously increasing with the variable, the discrepancy between probabilities, capturing the variation in uncertainty regarding the downturn may remain constant after some point as we have seen in this empirical example. The two approaches carry different information; on the one hand, the sample-based approach relies heavily on past behaviours and is usually more conservative, yielding higher probabilities of turning points. On the other hand the simulations-based approach yields probabilities solely induced by the underlying model. Both methods capture the bi-modality of potential future path characterising locally explosive episodes. They could, individually or combined, be used for investment strategies for instance. However, such strategy depends on what the investor is seeking, their risk aversion and their beliefs regarding the process.

6 Conclusion

This paper analyses and compares in details two approximation methods developed to forecast mixed causal-noncausal autoregressive processes. It focuses on $MAR(r,1)$ processes and aims attention at predictive densities rather than point forecasts as they are more informative, especially in the case of explosive episodes.

The sample-based (Gouriéroux and Jasiak, 2016) and simulations-based (Lanne, Luoto, and Saikkonen, 2012) methods are compared to theoretical results using various $MAR(0,1)$ processes and Cauchy-distributed errors. Results are also derived for Student's t distributions with 2 and 3 degrees of freedom even though no closed-form solutions exist. This paper focuses on one-step ahead forecasts to give a rigorous analysis of how and why estimations may differ from closed-form results. We find that theoretical and sample-based predictive densities start to differ as the series departs from central values, and the discrepancies increase with the level of the series, the lead coefficient and the sample size. The sample-based approach gives time-varying probabilities and depends on how similar the event under investigation is to past events. This approach yields results that are a mixture of probabilities ensuing from the underlying distribution and from past behaviours of the series. Simulations-based predictive probabilities are a good approximation of theoretical results obtained with Cauchy-distributed errors, as long as the number of simulations in

the approximations is sufficient and probabilities are also converging to a unique distribution for processes with $t(2)$ and $t(3)$ errors. Both methods capture the bi-modality of the conditional distribution as the series diverges from central values, which is an indicator of a potential bubble outset.

We illustrate the two methods with a detrended Nickel prices series. When the underlying distribution does not admit closed-form expressions for the predictive densities, the only way to determine the proportion of probabilities induced by past behaviours in the sample-based approach is to compare it with the probabilities obtained with the simulations-based method. This information can be used in investment decisions, where the strategy is to be built based on the investor's risk aversion and beliefs regarding the series.

References

Andrews, B., Davis, R., and Breidt, J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, *97*(7), 1638–1659.

Bec, F., Bohn Nielsen, H., and Saïdi, S. (2019). *Mixed causal-noncausal autoregressions: Bimodality issues in estimation and unit root testing* (Tech. Rep.).

Fries, S. (2018). Conditional moments of anticipative α-stable markov processes. *arXiv preprint arXiv:1805.05397*.

Fries, S., and Zakoïan, J.-M. (2019). Mixed causal-noncausal AR processes and the modelling of explosive bubbles. *Econometric Theory*, 1–37.

Gouriéroux, C., Hencic, A., and Jasiak, J. (2018). Forecast performance in noncausal MAR(1, 1) processes.

Gouriéroux, C., and Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, *37*(3), 405–430.

Gouriéroux, C., and Jasiak, J. (2018). Misspecification of noncausal order in autoregressive processes. *Journal of Econometrics*, *205*(1), 226–248.

Gouriéroux, C., Jasiak, J., and Monfort, A. (2016). Stationary bubble equilibria in rational expectation models. *CREST Working Paper. Paris, France: Centre de Recherche en Economie et Statistique*.

Gouriéroux, C., and Zakoïan, J.-M. (2013). Explosive bubble modelling by noncausal process. *CREST. Paris, France: Centre de Recherche en Economie et Statistique*.

Gouriéroux, C., and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(3), 737–756.

Hecq, A., Lieb, L., and Telg, S. (2016). Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*(123/124), 307–331.

Hecq, A., Lieb, L., and Telg, S. (2017). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The MARX package.

Hecq, A., Telg, S., and Lieb, L. (2017). Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, *5*(4), 48.

Hencic, A., and Gouriéroux, C. (2015). Noncausal autoregressive model in application to bitcoin/ exchange rates. In *Econometrics of risk* (pp. 17–40). Springer.

Karapanagiotidis, P. (2014). Dynamic modeling of commodity futures prices. *MPRA Paper 56805, University Library of Munich, Germany.*.

Lanne, M., Luoto, J., and Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, *28*(3), 623–631.

Lanne, M., Nyberg, H., and Saarinen, E. (2012). Does noncausality help in forecasting economic time series? *Economics Bulletin*.

Lanne, M., and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, *3*(3).

Lof, M., and Nyberg, H. (2017). Noncausality and the commodity currency hypothesis. *Energy Economics*, *65*, 424–433.